

Growth

Metric	Start	Now	Delta
Reviews	79,995	108,943	+28,948
DB size	45.9 MB	64.08 MB	+18.2 MB
Runs	0	13	—

~29k new reviews added over ~3 days across 11 full-20-app runs + 2 smoke tests.

Ingestion rates vary significantly by gap length

- **Run 3** (first full run): **7,336 new**: the big catch-up, since there was a time gap between the original scrape and now
- **Runs 7–9** (2-hour intervals on Feb 12): **1,380 → 883 → 653**: diminishing returns as the gap shrinks
- **Run 10** (after ~17h gap): **4,702**: jumps back up because reviews accumulated overnight
- **Runs 12–13** (4-hour intervals): **1,138 → 2,120**: steady-state for that cadence

This confirms the 4-hour interval is a sensible default, enough new reviews accumulate to make each run worthwhile.

Deduplication: Working perfectly

- **0 reviews linked to more than 1 run** in review_scrape_log. Every review is attributed to exactly the run that first discovered it. The pre-insert dedup check is doing its job.

Schema observations

Nullable fields are stable:

- app_version NULL: **14.4%** (was 14.1% in original data, consistent pattern, not a schema issue)
- reply_content NULL: **87.0%** (was 86.1%, devs reply to ~13% of reviews, stable ratio)
- content empty: **0**, every review has text

Rating distribution held steady: the new ingestion data matches the original bimodal pattern:

- 5-star dominance: 58.2% of new reviews (vs 58.6% original)
- 1-star: 26.1% of new (vs 24.5% original)

One friction point: TikTok

com.zhiliaoapp.musically started failing in runs 12–13, making those runs partial instead of completed. It already has 5,690 reviews in the DB from earlier successful runs. This is likely a transient Google Play API issue or rate-limiting. The pipeline handled it correctly: logged the error, continued with the other 19 apps, marked status as partial.

Summary

The schema is holding up well under continuous ingestion. No constraint violations, dedup is clean, nullable field ratios are stable, and the audit trail (scrape_runs + review_scrape_log) is fully populated. The only real observation is that the TikTok app ID intermittently fails, this is worth monitoring but not a schema issue.